A Decentralized Framework for AI-based Resource Allocation in Open RAN

Mrs. Gyara Monika *1 Assistant Professor, Department of CSE (Artificial Intelligence & Machine Learning), ACE Engineering College, Ankushapur, Hyderabad monika.gyara@aceec.ac.in (Corresponding Author)

Kosana HarshaVardhan *2 Student of ACE Engineering College, Department of CSE (Artificial Intelligence & Machine Learning) kosana.harshavardhan0705@gmail.com

Yash Vinay Chamle *3 Student of ACE Engineering College, Department of CSE (Artificial Intelligence& Machine Learning) chamleyash@gmail.com

Vangoori Harsha Vardhan *4 Student of ACE Engineering College, Department of CSE (Artificial Intelligence &Machine Learning)vangooriharsha17@gmail.com

Abstract: Modern mobile networks, particularly within the Open-RAN (O-RAN) architecture, increasingly rely on Artificial Intelligence for dynamic network slicing. However, conventional approaches often depend on a centralized AI model, which introduces significant challenges related to scalability, reliability as a single point of failure, and data privacy. This project proposes and implements a novel decentralized framework, to address these limitations using Federated Learning and Deep Learning. The system features autonomous DL agents, implemented in Pyt1orch, that reside at the network edge, making real-time resource allocation decisions based on local data. These agents collaboratively train a global intelligence model, orchestrated by a central FL server, by sharing only their learned model parameters, thus preserving data privacy. Through a custom-built simulation environment, this work demonstrates that the federated system successfully learns to manage network resources, achieving performance comparable to a centralized model while exhibiting graceful degradation and enhanced scalability.

Keywords: O-RAN, Federated Learning, Deep Learning, Resource Allocation, Decentralized AI, Edge Intelligence, Network Slicing, PyTorch, Privacy Preservation, Scalability.

1. INTRODUCTION

The rapid evolution of intelligent communication systems and edge computing has introduced new challenges in managing resources efficiently while ensuring scalability, low latency, and data privacy. Traditional centralized AI-based decision-making approaches rely heavily on collecting all operational data at a central server, which often leads to communication overhead, single points of failure, and increased privacy risks. These limitations become more pronounced in large-scale and real-time environments such as modern wireless networks, distributed systems, and edge-enabled infrastructures. As the number of connected devices and network nodes continues to grow, centralized learning frameworks struggle to adapt dynamically to changing conditions and localized requirements. Decentralized learning has therefore emerged as a promising alternative, enabling intelligent models to be trained closer to the data source while reducing dependence on centralized control. Federated learning, in particular, allows multiple edge nodes to collaboratively learn a global model without sharing raw data, thereby preserving privacy and reducing communication costs. However, implementing federated intelligence in practical systems introduces challenges related to coordination, convergence efficiency, hardware constraints, and performance evaluation. Many existing studies focus primarily on software simulations and overlook real-world deployment considerations. To address these challenges, this project proposes a decentralized, federated learning framework integrated with edge hardware to enable realistic and scalable intelligent resource allocation. By combining local model training, secure parameter aggregation, and iterative global updates, the system aims to achieve robust performance while maintaining adaptability to dynamic environments. This approach bridges the gap between theoretical federated learning models and practical hardware-based implementations, making it suitable for next-generation intelligent networked systems.

2.PROPOSED METHODOLOGY

The The proposed decentralized learning framework is designed to enable intelligent resource allocation through collaborative training across multiple edge nodes while preserving data privacy and reducing centralized dependency. Each edge node operates independently, collecting local network or system data and training a lightweight AI model tailored to its environment. These locally trained models do not share raw data; instead, only learned parameters are transmitted to a federated aggregation server. The central server performs secure parameter aggregation using a federated averaging mechanism to generate a global model. This global model is redistributed to all participating edge nodes for further local refinement. The framework supports iterative training cycles, ensuring adaptability to dynamic conditions. Performance metrics such as latency, convergence speed, and accuracy are monitored to evaluate effectiveness. By combining decentralized learning, edge intelligence, and controlled aggregation, the methodology ensures scalability, robustness, and real-world deployability in hardware-constrained environments.

3.LITERATURE SURVEY

[1] Title: Slice-on-the-Fly: AI-based Network Slicing in O-RAN for Dynamic Traffic Demands

Authors: A. Alchaab, A. Younis and D. Pompili

This This paper addresses the challenge of managing diverse, dynamic traffic in Open Radio Access Networks (O-RAN) to guarantee Quality of Service (QoS). The authors formulate resource allocation as a complex Nonlinear Programming problem, which is impractical for real-time adjustments. To solve this, they decompose the issue into Long-Term Resource Allocation (LTRA) and Short-Term Resource Scheduling (STRS). They introduce a novel AI algorithm, LS RLSlice, which combines Long Short-Term Memory (LSTM) with Deep Deterministic Policy Gradient (DDPG) reinforcement learning. This approach allows the network to adaptively allocate resources "on-the-fly." The solution is validated through simulations, where it outperforms existing methods, and on the real-world POWDER testbed, demonstrating its practical applicability within the O-RAN RIC framework.

[2] Title: Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges

Author: Michele Polese, Leonardo Bonati, Salvatore D'Oro, Stefano Basagni, and Tommaso Melodia.

This comprehensive tutorial explains the Open Radio Access Network (O-RAN) architecture, built on principles of disaggregation, intelligence via RAN Intelligent Controllers (RICs), virtualization, and open interfaces. It details the Near-RT and Non-RT RICs, key interfaces like E2, A1, O1, and Fronthaul, and the AI/ML workflow O-RAN enables. The paper discusses use cases, security challenges arising from the open architecture, standardization efforts, and experimental platforms. It concludes by highlighting research gaps, emphasizing O-RAN's potential to revolutionize cellular networks through enhanced flexibility, innovation, and data driven control.

[3] Title: ORANSlice: An Open-Source 5G Network Slicing Platform for O-RAN

Authors: Hai Cheng, Salvatore D'Oro, Rajeev Gangula, Sakthivel Velumani, Davide Villa, Leonardo Bonati, Michele Polese, Gabriel Arrobo, Christian Maciocco, and Tommaso Melodia.

This paper introduces ORANSlice, an open-source platform designed to implement 5G network slicing within the O-RAN framework, addressing the lack of such tools in the community. Built upon the OpenAirInterface (OAI) stack, it features a 3GPP-compliant, two-tier scheduler for managing RAN slice resources and extends the OAI UE to support multiple PDU sessions, enabling multi slice capabilities on a single device. ORANSlice integrates an E2SM-CCC service model and an associated xApp for controlling slicing via the Near-RT RIC. The platform's effectiveness was validated on different O-RAN testbeds, demonstrating slice prioritization and minimum resource guarantees.

[4] Title: Network Slicing Meets Artificial Intelligence: An AI-Based Framework for Slice Management

Authors: Dario Bega, Marco Gramaglia, Andres Garcia-Saavedra, Marco Fiore, Albert Banchs, and Xavier.

This article proposes a comprehensive framework leveraging Artificial Intelligence (AI) to manage network slices throughout their lifecycle. It breaks down slice management into three key functions operating at different timescales: AI-based admission control (hours) using DRL to maximize operator revenue; AI based resource orchestration (minutes) using CNNs to forecast capacity and minimize operational costs; and AI-based radio scheduling (sub-seconds) using unsupervised learning and DRL to handle complex RAN resource allocation. Case studies show significant performance gains (25-80%), demonstrating AI's potential for automating complex slice management tasks.

[5] Title: Deep Learning for Intelligent and Automated Network Slicing in 5G Open RAN (ORAN)Deployment
Authors: Shu-Ping Yeh, Sonia Bhattacharya, Rashika Sharma, and Hassnaa Moustafa.

This paper presents an intelligent xApp for automated RAN slicing within the O-RAN architecture. The xApp integrates AI, specifically deep learning models like LSTM, Seq2Seq, or TCN, for traffic load prediction, coupled with a radio resource planning module. It dynamically calculates and configures 3GPP-compliant Radio Resource Management (RRM) policies (dedicated, minimum, maximum ratios) based on predicted load and SLA requirements. The xApp was implemented and evaluated using an O-RAN compliant Near-RT RIC (ONF SD RAN) and an E2 simulator, demonstrating its ability to automate slice management intelligently and efficiently, particularly improving delay performance for sensitive slices.

[6] Title: Open RAN LSTM Traffic Prediction and Slice Management using Deep Reinforcement Learning

Authors: Fatemeh Lotfi and Fatemeh Afghah.

The study proposes an integrated framework that combines traffic prediction with dynamic network slice management to enhance resource allocation efficiency. Specifically, an LSTM-based rApp deployed in the non-real-time RIC forecasts traffic patterns, and these predictions are leveraged by a DDRL-based xApp in the near-real-time RIC to optimize slice resource allocation. By incorporating traffic awareness into the decision-making process, the framework enables more proactive and adaptive control compared to conventional DRL approaches. Experimental results demonstrate that the integration of LSTM-driven traffic forecasting with DDRL significantly improves system performance, achieving up to 7.7% higher cumulative return while also reducing SLA violations relative to DRL models that operate without traffic prediction.

4. SYSTEM ARCHITECTURE

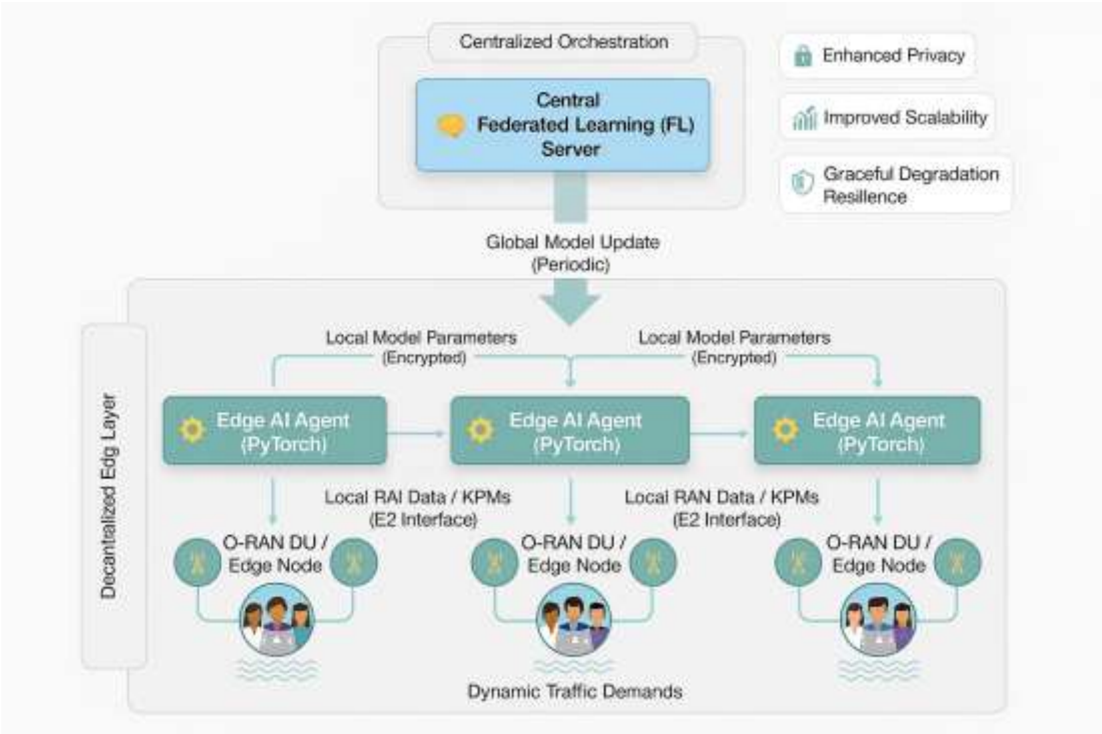

Figure 1 : System Architecture Diagram

S. No	Author Name	Title	Methodology	Findings
1	Adhwaa Alchaab; Ayman Younis; Dario Pompili	Slice-on-the-Fly: AI-based Network Slicing in O-RAN for Dynamic Traffic Demands	A decentralized framework was implemented where numerous Deep Reinforcement Learning agents at the network edge are collaboratively trained using a Federated Learning server.	The federated system successfully learns to manage network slices with performance comparable to a centralized model, while providing superior scalability, fault tolerance, and data privacy.
2	Michele Polese, Leonardo Bonati, Salvatore D'Oro, Stefano Basagni, Tommaso Melodia	Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges	The paper presents the O-RAN architecture, including RICs, O-Cloud, open interfaces and AI/ML workflows for network control.	This shows that O-RAN's design promotes innovation and reduces vendor lock-in, while highlighting challenges in interoperability, security and multi-time-scale AI algorithms.
3	Hai Cheng, Salvatore D'Oro, Rajeev Gangula, Sakthivel Velumani, Davide Villa, Leonardo Bonati, Michele Polese, and Tommaso Melodia.	ORANSlice: An Open-Source 5G Network Slicing Platform for O-RAN	This paper surveys ORANSlice, an open-source O-RAN slicing platform with a two-tier scheduler, multi-slice support, and O-RAN-compliant control.	Studies ORANSlice performs end-to-end O-RAN slicing on testbeds, enforcing policies for minimum resource allocation and slice prioritization.
4	Dario Bega, Marco Gramaglia, Andres Garcia-Saavedra, Marco Fiore, Albert Banchs, and Xavier Costa-Perez.	Network Slicing Meets Artificial Intelligence: An AI-Based Framework for Slice Management	This paper presents an AI-based framework managing network slices across timescales: DRL for hours, CNN for minutes, and unsupervised learning with DRL for seconds.	The study shows AI-based slice management improves performance by 25–80%, reduces costs by over 50%, and achieves up to 30% CPU savings compared to traditional schedulers.
5	Shu-Ping Yeh, Sonia Bhattacharya, Rashika Sharma, and Hassnaa Moustafa.	Deep Learning for Intelligent and Automated Network Slicing in 5G Open RAN (ORAN) Deployment	An intelligent NSRRM xApp automates RAN slicing, leveraging LSTM, Seq2Seq, and TCN models, integrated with near-RT RIC.	This framework enables automated, intelligent network slicing, improving delay for latency-sensitive slices and calculating PRBs while controlling the RAN O-RAN-compliantly.
6	Fatemeh Lofti, Fatemeh Afghan.	Open RAN LSTM Traffic Prediction and Slice Management using Deep Reinforcement Learning	It uses a framework that integrates traffic prediction with dynamic slice management, using LSTM in non-RT RIC to guide DDRL in near-RT RIC for resource allocation.	The study shows that combining the LSTM rApp with DDRL xApp improves performance, achieving up to 7.7% higher cumulative return and reducing SLA violations compared to DRL without traffic forecasts.

6. REFERENCES :

[1] ShiA. Alchaab, A. Younis, and D. Pompili, "Slice-on-the-Fly: AI-based Network Slicing in O-RAN for Dynamic Traffic Demands," in *Proc. 2025 IEEE 26th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, Fort Worth, TX, USA, 2025, pp. 51-60, doi: 10.1109/WoWMoM65615.2025.00020.

[2] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1376-1411, Second Quarter 2023, doi: 10.1109/COMST.2023.3239220.

[3] H. Cheng et al., "ORANSlice: An Open-Source 5G Network Slicing Platform for O-RAN," in Proc. 30th Annual International Conference on Mobile Computing and Networking (ACM MobiCom '24), Washington D.C., DC, USA, Nov. 2024, pp. 1-6, doi: 10.1145/3636534.3701544. Buse, Raymond P. L., and Weimer, Westley (2010). Learning a Metric for Code Readability. This work introduces a readability metric based on human judgment and code structure.

[4] D. Bega et al., "Network Slicing Meets Artificial Intelligence: An AI-Based Framework for Slice Management," IEEE Communications Magazine, vol. 58, no. 6, pp. 32-38, June 2020, doi: 10.1109/MCOM.001.1900653.

[5] SHU-PING Yeh, S. Bhattacharya, R. Sharma, and H. Moustafa, "Deep Learning for Intelligent and Automated Network Slicing in 5G Open RAN (ORAN) Deployment," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 64-70, 2024, doi: 10.1109/OJCOMS.2023.3337854.

[6] F. Lotfi and F. Afghah, "Open RAN LSTM Traffic Prediction and Slice Management using Deep Reinforcement Learning," *arXiv preprint arXiv:2401.06922*, Jan. 2024. arxiv.org/abs/2401.06922